

DATA PLOTTING EXTENSION FOR STRUCTURED QUERY LANGUAGE

DESCRIPTION

Background of Invention

[Para 1] This invention relates to queries directed to relational databases using Structured Query Language (SQL), and more particularly to obtaining graphical representations (data plots) of the data retrieved in accordance with those queries.

[Para 2] Structured Query Language (SQL) is an industry standard for writing queries to retrieve data from relational databases. In a typical database system (often called a relational database management system or RDBMS), shown schematically in Figure 1, database 1 is controlled by database software on a server device 2. A user enters a query, written in SQL, on an input device 3 connected to a client device 4; the transmitted query 5 causes a dataset 6 to be returned from the database. In general, execution of an SQL query by a database system returns a text based dataset. A typical query is shown in Figure 2A, requesting a count of selected distinct part numbers. As shown in Figure 2B, the result of the query is a list 21 of part numbers and the corresponding counts 22, presented in two columns as text data (usually in a standard ASCII format). Usually the user wishes to process the data further, for example by making input to a calculation or presenting the data in the form of a graph.

[Para 3] If a graphical presentation of the text data is desired, the user must then manipulate the text data and use plotting software or plot the data manually. In either case, the user must perform a number of additional steps to obtain the data as a graph instead of text; furthermore, these steps are not part of the query process. At present there is no interface defined in the SQL standard which allows a user to specify that the query results should be

plotted in a graph instead of being returned as raw text data. In particular, there are no existing keywords or user-selectable response formatting syntax which would permit graphical display of the returned query results.

[Para 4] Several systems and techniques have been proposed for generating database queries so that the queries may be viewed in graphical form. For example, "User-Defined Visual Query Language" (Cruz, Proc. IEEE Symposium on Visual Languages, St. Louis, Oct. 1994, p.224) describes an object-oriented database query language in which objects are given graphical definitions, so that the structure of the query may be viewed as a picture on a screen, and a user can extract information about the data in a visual fashion. Accordingly, the query is presented in a visualized form. U.S. Pat. No. 6,578,028 (Egilsson et al.) describes a graphical SQL query generator by which a standard SQL query may be constructed using a graphical user interface (GUI). Several statistical methods are described for calculating dependency matrices among the data in the database; the dependency matrices are associated with a relation definition. A GUI may be used with the matrices to construct an SQL query.

[Para 5] In addition, systems have been developed wherein types and characteristics of data returned from a database are presented graphically. U.S. Pat. No. 6,014,661 (Ahlberg et al.) describes an automated method for analyzing data fields in a database, in which data is downloaded from a database and relationships within the data set are displayed graphically; a GUI is used for interactive exploration of the database. Although this system facilitates analysis of and browsing in the database, the returned result of a query is still formatted only as text data. U.S. Pat. No. 5,636,350 (Eick et al.) describes a system in which characteristics of a result of a query are displayed graphically, and symbols are displayed whose appearance varies according to the number of items of data satisfying the query (number of hits).

[Para 6] None of the above-noted references discusses a method for plotting the actual data returned by a database query (as opposed to characteristics of the data). Furthermore, these systems do not include a way to specify graph plotting as part of the SQL query. There remains a need for a method which

allows a user to query a database and graphically view the results without any intermediate steps.

Summary of Invention

[Para 7] The present invention addresses the above-described need by providing an extension to the SQL standard for plotting the results of a query. In particular, the invention provides keywords for specifying a format for graphing selected data, and syntax for recognizing those keywords, thereby causing the data to be presented as a graph according to the specified format. This extension of SQL maintains the syntax and style of conventional SQL queries. This permits automated systems, such as database driven websites, to issue extended SQL queries directly to a relational database and have the results returned as formatted graphical content.

[Para 8] In accordance with a first aspect of the invention, a method is provided for obtaining information from a relational database. This method includes formulating a query to retrieve data from the database; a first portion of the query specifies the data to be retrieved, and a second portion of the query specifies a format for graphing the data. The query is then transmitted to the database, and the data is returned from the database in accordance with the query. The data is then presented in accordance with the specified format. The first portion of the query and the second portion of the query are preferably formulated in SQL. The data may be returned as a binary image, or alternatively as an image representation of the data in ASCII format. As part of the step of returning the data, the query is preferably interpreted in accordance with SQL having keywords and syntax for specifying the format (that is, instructions for presenting the data as a graph). The graphical image may be for example a line graph, a horizontal bar chart, a vertical bar chart, a pie chart, a scatter plot, a contour plot, or a wafer map, in accordance with a keyword in the second portion of the query.

[Para 9] Furthermore, the process of returning the data may include interpreting the first portion of the query to cause the data to be retrieved from the database; creating a dataset for the data; incorporating the data into

the dataset; and constructing a graphical image using the data, in accordance with the specified format. The step of interpreting may further include parsing the query so that the first portion of the query and the second portion of the query are interpreted separately.

[Para 10] In accordance with another aspect of the invention, a method is provided for querying a database which includes the steps of specifying the data to be returned from the database in a first portion of a query, and specifying a format for graphing the data in a second portion of the query. The data are thus returned as a graphical image in accordance with the specified format. The first portion and second portion of the query are both preferably in a structured query language (SQL) which includes keywords and syntax for specifying the format. The graphical image may be either a binary image, or a representation of the data in ASCII format; the graphical image may be a line graph, a horizontal bar chart, a vertical bar chart, a pie chart, a scatter plot, a contour plot, or a wafer map, in accordance with a keyword in the second portion of the query.

[Para 11] In accordance with an additional aspect of the invention, a system is provided for retrieving and presenting data from a database. This system includes the database; an input device for entering a database query; a device for interpreting the query which is effective to format the data for presentation in graphical form; and an output device for presenting the data as a graphical image. The query preferably is formulated in a structured query language (SQL), and includes a first portion specifying the data to be retrieved and a second portion specifying the graphical form.

[Para 12] In this system, the device for interpreting the query preferably has an interpreter for interpreting both the first portion and the second portion of the query. Alternatively, the device for interpreting the query may have a first interpreter for interpreting the first portion of the query and a second interpreter for interpreting the second portion of the query. In the latter case the first interpreter is effective to cause return of the data from the database in accordance with the first portion of the query, while the second interpreter is effective to parse the query into the first portion and the second portion;

create a dataset for the data; incorporate the data into the dataset; and construct the graphical image using the data, in accordance with the second portion of the query.

[Para 13] According to a further aspect of the invention, a computer-readable storage medium is provided which has stored therein instructions for performing a method for obtaining information from a relational database, as described above. This method includes querying a database to retrieve data therefrom; transmitting the query to the database; causing the data to be returned from the database in accordance with the query; and presenting the data in accordance with a specified format.

[Para 14] According to another aspect of the invention, a computer program product is provided for performing a method as described above.

[Para 15] According to a further aspect of the invention, an improvement of a computer program product for interpreting a structured query language is provided. This improvement includes first computer program code for recognizing keywords for specifying a format for graphing data returned from a database, and second computer program code for causing the data to be presented as a graph according to the specified format. The first computer program code may include code for recognizing a delimiter keyword which separates SQL statements in a first portion of the query specifying the data from SQL statements in a second portion of the query specifying the format.

Brief Description of Drawings

[Para 16] Figure 1 schematically illustrates a typical database system in which an SQL query is made and data is returned from a database.

[Para 17] Figure 2A shows a typical SQL query.

[Para 18] Figure 2B shows the format of the result of the query of Figure 2A.

[Para 19] Figure 3 shows the syntax used to generate a plot of data from an SQL query, in accordance with the present invention.

[Para 20] Figure 4A shows the SQL query of Figure 2A, modified in accordance with the invention to generate a graph of the data returned from the database.

[Para 21] Figure 4B shows the plot generated from the query of Figure 4A.

[Para 22] Figure 4C shows an example of a plot in ASCII format.

[Para 23] Figure 5A shows another example of an SQL query according to the present invention.

[Para 24] Figure 5B shows the plot generated from the query of Figure 5A.

[Para 25] Figure 6 schematically illustrates a relational database including an SQL interpreter which receives SQL queries.

[Para 26] Figure 7A illustrates an embodiment of the invention in which the SQL interpreter receives SQL queries formulated using the syntax of the SQL extension according to the present invention.

[Para 27] Figure 7B is a flowchart showing steps carried out by the SQL interpreter of Figure 7A.

[Para 28] Figure 8A illustrates another embodiment of the invention in which a standard SQL interpreter communicates with an additional SQL interpreter capable of receiving SQL queries formulated using the syntax of the SQL extension according to the present invention.

[Para 29] Figure 8B is a flowchart showing steps carried out by the additional SQL interpreter of Figure 8A.

Detailed Description

[Para 30] SQL Extension: Keywords and Syntax

[Para 31] In a preferred embodiment, the SQL standard is extended by adding the keywords listed below in Table 1.

Table 1: Keywords for SQL Data Plotting Extension

PLOT	VAR	ON	
XAXIS	Y1AXIS	Y2AXIS	ZAXIS
TITLE1	TITLE2	LEGEND	
FOOTNOTE1	FOOTNOTE2		
LINE	HBAR	VBAR	
PIE	SCATTER	CONTOUR	WMAP
FILE	IMAGETYPE		
XSIZE	YSIZE		

[Para 32] The syntax used to generate a plot from a basic SQL query is illustrated in Figure 3 and described in more detail below. In Figure 3, bracketed statements are optional. The PLOT statement 31 is used to specify that the data from the query is to be plotted. (It will be appreciated that standard SQL "helper" words, e.g. BY, AS, THE, are used freely.) The data is obtained by constructing a query generally including SELECT and FROM, as is known in the art. The variables to be plotted, specified using the VAR keyword 32, may be any of the columns of data in the dataset (that is, any selected column 33 in the text format of the dataset, indicated by <column> in Figure 3). The axes 34 to be used in the plot (XAXIS, Y1AXIS, Y2AXIS or ZAXIS) are specified using the ON keyword 35. Each of the axes may be given a label 36, illustrated in Figure 3 as <label>. The plot may also be given titles and footnotes using the keywords TITLE1, TITLE2, FOOTNOTE1, FOOTNOTE2 37–40; the text of the titles 41 and footnotes 42 are illustrated in Figure 3 as <titletext> and <footertext> respectively. If it is desired to save the plot as a file with a selected image type, a file name <filename> is assigned using the FILE keyword 43. An image type <imagetype> is assigned using the IMAGETYPE keyword 44; otherwise the default image type is used. The size of the displayed plot in the x and y directions may be specified using the keywords XSIZE 45 and YSIZE 46, respectively. In addition, the location of a

legend indicating the plotted variables may be specified using the optional LEGEND keyword 47.

[Para 33] It will be appreciated that different names could be chosen for the keywords enabling the functions described herein.

[Para 34] A number of different types of plots may be drawn from the same data; that is, <charttype> in Figure 3 may be replaced by any of these keywords: LINE, HBAR, VBAR, PIE, SCATTER, CONTOUR, and WMAP. The various types of plots are therefore as shown in Table 2.

Table 2: Types of Plots for SQL Data Plotting Extension

LINE	Line graph
HBAR	Horizontal bar graph
VBAR	Vertical bar graph
PIE	Pie chart
SCATTER	Scatter plot
CONTOUR	Contour plot
WMAP	Wafer map

[Para 35] It will be appreciated that still other types of charts may be defined and plotted, in addition to those discussed herein.

[Para 36] An example of an application of this syntax is shown in Figures 4A and 4B. The first three lines 51 of this query are identical to the query shown in Figure 2B. The last four lines 52 of the query begin with PLOT AS VBAR, meaning that the returned data is to be plotted as a vertical bar graph. The keyword VAR introduces the variables to be plotted; the partnumber 21 on the x-axis, and the total 22 on the y-axis (compare Figure 2B). The labels for these axes (each following AS) are specified as PARTNUMBER and TOTAL respectively. The title of the graph is specified using the TITLE1 keyword, and

a footnote is specified using the FOOTNOTE1 keyword. The LEGEND keyword is not used in the query, so a legend is placed in a default location to the right of the graph.

[Para 37] It should be noted that the query of Figure 4A returns pixel data in a binary image format as shown in Figure 4B. The data is thus presented graphically as a direct result of the query, without intermediate steps (such as invoking another programming application to manipulate the raw returned data). A comparison of Figure 4B with Figure 2B shows that the graph of Figure 4B presents the same data as the raw data of Figure 2B.

[Para 38] Alternatively, the query may simply return a representation of the data in ASCII format. An example of a plot in ASCII format is shown in Figure 4C.

[Para 39] Another example of the syntax of the SQL extension of the invention is shown in Figures 5A and 5B. This example shows distinct sets of data presented on the same graph. The SELECT statement 61 specifies that columns of data labeled chipx, chipy, chipx2 and chipy2 are to be returned from the database. The PLOT statement 62 specifies that a scatter plot is to be produced with the variable chipx plotted in the x-direction, and two variables chipy and chipx2 plotted in the y-direction. The variables chipx, chipy and chipx2 are given the labels "chip x", "chip y" and "y2chip". (Note that in this example not all of the returned data is included in the plot.) According to this syntax, in the PLOT statement VAR is followed by the variable name, ON by the axis on which that variable is to be plotted, and AS by the label given that variable on the graph. The x-axis, left-hand y-axis, and right-hand y-axis of the graph are given the labels for the XAXIS, Y1AXIS and Y2AXIS variables respectively. Although each variable name (e.g. chipx2) must match a column of data in the dataset, this group of data may be plotted with a different label (in this case, "y2chip").

[Para 40] Implementation of Extended SQL

[Para 41] A database 1 implementing the present invention is shown schematically in Figure 6. Database 1 is controlled by a database management system and includes a data storage medium 12 and an SQL interpreter 10. An

incoming query 5, written in SQL, causes the SQL interpreter to send an internal request 11 for return of data 13. (This internal request is in general not made in SQL, but is a set of machine-readable instructions.) In a preferred embodiment of the invention, the SQL interpreter 10 is capable of interpreting the extended SQL queries described above, so that the output 65 from the database 1 is a binary image which can be presented as a graph. A system including a database with such an interpreter is shown schematically in Figure 7A. The query generator 71 (e.g. a user making input to a client system) transmits query 5 in the extended SQL; the output 65 is a binary image which may be in any of a variety of formats, but is generally binary pixel data in a standard image format such as .jpeg, .gif, .png or the like. The output 65 may thus be viewed as a graphical image 72 on any of a number of media, including a display screen, a web page, a desktop application, etc.; alternatively, the image may be stored as a file on a data storage medium.

[Para 42] Figure 7B is a flowchart showing the sequence of steps executed by the interpreter 10 in this embodiment. The interpreter accepts a query transmitted from the input device 71 (step 701). In step 702, the interpreter determines whether the query is valid (that is, whether all of the keywords are recognized and the statements have proper syntax). An error message is returned if the query is invalid. A valid query is separated into a first portion in "regular" or standard SQL and a second portion in extended SQL (step 703). The keyword PLOT may be used as a convenient delimiter between the first and second portions of the query. The portion of the query in standard SQL is executed (step 704), resulting in data being returned from the database. If a plot of the data is requested as part of the query (step 705), the interpreter proceeds to create the plot; otherwise the raw data is returned to the output device 72 (step 706). In step 707, the plot is created based on the arguments to the PLOT statement. The interpreter searches the PLOT statement for the plot specifications, and prepares those specifications as inputs to a charting routine. (The charting routine may be integrated with the interpreter, but is typically a separate third-party program.) The charting routine draws the plot in the specified format and inserts the data. The completed plot (a binary

image as described above) is then transmitted to the output device 72 (step 708).

[Para 43] If the interpreter resident in the database is not capable of recognizing the SQL extensions of the present invention, the invention may be implemented using the system shown schematically in Figure 8A. In this embodiment of the invention, the database interpreter 15 only recognizes the conventional SQL; an additional interpreter 16 is used to parse the query 5 into recognizable and non-recognizable tokens, with the PLOT keyword serving as a delimiter. Interpreter 16 is itself a program, written in a general-purpose programming language such as PHP, Java or the like, and may advantageously include conventional charting software. The standard portion of the SQL query is passed on to the database 1. The interpreter 16 creates a dataset containing only those data that will be plotted, in accordance with the PLOT statement. When data 6 returns from the database (raw data in ASCII format), that data is incorporated into the dataset and passed to the charting software, which then constructs a graphical image using the dataset as specified in the query. The output 65 of interpreter 16 is then a graphical image as specified in the PLOT statement in the query. Interpreter 16 may be a stand-alone program as shown in Figure 8A, or may be integrated into the RDBMS.

[Para 44] Figure 8B is a flowchart showing the sequence of steps executed by the interpreter 16 in this embodiment. In this example, the interpreter 16 is assumed to be separate from the database. The interpreter accepts a query, determines whether it is valid, and separates the query into a portion in "regular" SQL and a portion in extended SQL, as in the previous embodiment (steps 801–803). The interpreter then establishes a connection to the database (step 804), and sends the first portion of the query to the database as a standard SQL query (step 805). In response to this query, the interpreter receives raw data from the database (step 806); the interpreter then disconnects from the database (step 807). If a plot is not requested as part of the original query (step 808), the raw data is returned to the output device 72 (step 809). If a plot is requested, the interpreter proceeds to create the plot in accordance with the PLOT statement in the query (step 810). The interpreter

16 may include a charting routine, or alternatively may send the plot specifications to a separate charting routine. The completed plot is then returned to the output device 72 as a binary image (step 811).

[Para 45] It should be noted that from the viewpoint of the user, the systems shown in Figure 7A and Figure 8A are equivalent. In both cases, the user need specify the graphical format for viewing the data only once, as an integral part of the query.

[Para 46] The above-described extension to SQL has been implemented on a system having an IBM DB2 database. Specifically, the features of the SQL extension were coded in PHP as a wrapper around the database. This implementation was made as shown in Figures 8A and 8B. However, as discussed above, it is preferable for the database to support the SQL extension directly, as an addition to the existing SQL executor.

[Para 47] A distinction should be drawn between retrieving a plot from a database and plotting data retrieved in response to a query. In the foregoing descriptions, an image representation of raw ASCII data is returned as a result of a query having a PLOT statement. The database may also be configured to include binary long objects (BLOB objects). As is understood by those skilled in the art, a BLOB is a database table column type, used to store binary data in a database. Such binary data may be retrieved directly from the database using standard SQL. Since a BLOB is capable of storing any binary content, it is possible for a BLOB to contain a binary image such as a plot. However, receiving BLOB data using a standard SQL query is different from making a query in the extended SQL of the invention, wherein a PLOT statement specifies how the data is to be presented as a binary image. In the latter case, the data selected in the first part of the query (i.e. with the SELECT statement) is not returned to the user directly. Instead, the interpreter creates a new representation of the original data according to the options specified by the user in the PLOT statement. In contrast, a query in standard SQL involving BLOB data results in the database returning the original data requested in the SELECT statement, without any additional interpretation. Furthermore, binary

image data (BLOB objects) in general cannot be plotted, and thus would not be recognized by an SQL query made according to the present invention.

[Para 48] While the invention has been described in terms of specific embodiments, it is evident in view of the foregoing description that numerous alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, the invention is intended to encompass all such alternatives, modifications and variations which fall within the scope and spirit of the invention and the following claims.

[Para 49] I claim: